

THE MULTIPLE COMPARISON PROCEDURES

by

NANCY CHIANG

B. A. NATIONAL TAIWAN UNIVERSITY, 1958

A MASTER REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

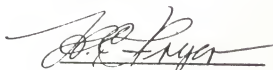
Department of Statistics

Kansas State University

Manhattan, Kansas

1967

Approved by

A handwritten signature in dark ink, appearing to read "D. E. Fryer", is written over a horizontal line.

Major Professor

21
2668
R4
1967
2532
B.2

TABLE OF CONTENTS

I	INTRODUCTION	1
II	GENERAL ASSUMPTIONS AND DECISIONS	2
III	FISHER'S LEAST SIGNIFICANT DIFFERENCE TEST (LSD)	4
IV	STUDENT-NEWMAN-KEULS MULTIPLE RANGE PROCEDURE (NMR)	5
V	DUNCAN'S NEW MULTIPLE RANGE TEST (DMR)	7
VI	TUKEY'S SIGNIFICANT DIFFERENCE RULE (TSD) BASED ON ALLOWANCE	9
VII	SCHEFFÉ'S SIGNIFICANT DIFFERENCE RULE BASED ON CONTRAST (SSD)	10
VIII	EXPMPL E AND COMMENTS	14
IX	BAYESIAN APPROACH TO MULTIPLE COMPARISONS	22
X	MONTE CARLO STUDIES OF FISHER'S LSD	32
XI	ACKNOWLEDGEMENT	37
XII	REFERENCE	38

I Introduction

The common practice for testing the homogeneity of a set of n treatment means in an analysis of variance is to use the F test, where $F = (\text{treatment mean square}) / (\text{error mean square})$. The procedure has special desirable properties for testing the homogeneity hypothesis that the n population treatment means (μ_i) are equal. When the resulting F is larger than the corresponding tabulated F at a previously chosen level of significance α , it is concluded that there are some difference among the n treatment means. However, this test gives no decision as to which of the μ_i may be considered different and which may not. Hence several test procedures have been proposed for solving this problem.

One of the simplest test procedures is R. A. Fisher's (5) least significant difference test. Some others are Student-Newman-Kuels (7,8,12) multiple range procedure, D. B. Duncan (1,3) new multiple range test, J. B. Tukey (13) significant difference rule and H. Scheffé (11) significant difference rule. Recently, Duncan (2,3) suggested a Bayesian least significant difference test for a symmetric multiple comparison of n treatment means.

The purpose of this report is to discuss the various test procedures and the basic points of difference among them; a complete set of results from digital computer to illustrate the protection and the power of the Fisher LSD test procedure also is presented.

II General Assumptions and Decisions

In the general problem we are given a sample of n observed means, say $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$, which are assumed to be drawn independently from n normal populations with true means $\mu_1, \mu_2, \dots, \mu_n$, respectively, and a common unknown variance σ^2 . s^2 , the error mean square is an unbiased estimate of σ^2 and is independent of the observed sample means, based on the degrees of freedom f .

In the simplest case with $n = 2$, only two sample means \bar{x}_1 and \bar{x}_2 , there are three possible decisions:

- (1) \bar{x}_1 is significantly less than \bar{x}_2 ,
- (2) \bar{x}_2 is significantly less than \bar{x}_1 ,
- (3) \bar{x}_1 and \bar{x}_2 are not significantly different.

When $n = 3$, there are 19 possible decisions:

- (1) Six decisions of the form: $\bar{x}_1 < \bar{x}_2 < \bar{x}_3$ (1,2,3)
i.e. (1,2,3), (1,3,2), (2,1,3), (2,3,1), (3,1,2) and (3,2,1).
- (2) Three decisions of the form: $\bar{x}_1 = \bar{x}_2$ but $\bar{x}_1, \bar{x}_2 < \bar{x}_3$ (1,2, 3)
i.e. (1,2, 3), (1,3, 2) and (2,3, 1).
- (3) Three decisions of the form: $\bar{x}_1 < \bar{x}_2$ but $\bar{x}_2 = \bar{x}_3$ (1, 2,3)
i.e. (1, 2,3), (2, 1,3) and (3, 1,2).
- (4) Six decisions of the form: $\bar{x}_1 < \bar{x}_3$ but $\bar{x}_1 = \bar{x}_2$, $\bar{x}_2 = \bar{x}_3$ (1,2,3)
i.e. (1,2,3), (1,3,2), (2,1,3), (2,3,1), (3,1,2) and (3,2,1).
- (5) One decision of the form: $\bar{x}_1 = \bar{x}_2 = \bar{x}_3$ denoted as (1,2,3).

The number of decisions increases very rapidly as n increases, making the situation very complicated. Following are just some of the examples.

- (1) $n!$ decisions of the form (1,2,...,n) with no underscoring.
- (2) $n!(n-1) / 2$ decisions of the form (1,2,3,...,n) with one pair of means underscored.

- (3) $n!(n-2) / 3$ decisions of the form (1,2,3,4,...,n) with three means underscored.
- (4) $n!(n-2)$ decisions of the form (1,2,3,4,...,n) with two pairs of means overlapped.
- (5) One decision of the form (1,2,...,n) with all means underscored, which implies that all n means are not significantly different from one another.

If one wishes to know which treatment means are equal and which are not, no matter whether they are greater (less) than the others, the decision problem can be simplified. In general, we have $\binom{n}{2}$ for two treatment-mean comparison, $\binom{n}{3}$ for three-mean comparison, $\binom{n}{4}$ for four-mean comparison,..., $\binom{n}{n-1}$ for $(n-1)$ -mean comparison provided that the prior F -test for testing the equality of n treatment means is significant.

III Fisher's Least Significant Difference Test (LSD)

If an analysis of variance leads to rejection of the hypothesis that n treatment means are equal, given equal variance at a predetermined level of significance α , one uses the t-test on all $\left(\frac{n}{2} \right)$ possible hypotheses that $\mu_i = \mu_j$ for all $i \neq j$ with the same α level of rejection. Then Fisher's (5) least significant difference LSD is defined as:

$$LSD_{\alpha} = t_{\alpha, f \text{ d.f.}} \sqrt{s^2 \left(\frac{1}{r_i} + \frac{1}{r_j} \right)} \quad (3.1)$$

where s^2 is the error mean squares in the preceeding ANOVA,

f is the error degrees of freedom associated with s^2 ,

r_i, r_j are the number of replicates in the $i^{\text{th}}, j^{\text{th}}$ treatment.

When $r_i = r_j = r$,

$$LSD_{\alpha} = t_{\alpha, f \text{ d.f.}} \sqrt{2s^2 / r} = t_{\alpha, f \text{ d.f.}} s_{\bar{x}} \sqrt{2} \quad (3.2)$$

where $s_{\bar{x}}$ is the standard error of mean.

Each difference between means is compared with the LSD_{α} . If the difference $\bar{x}_i - \bar{x}_j$ exceeds the LSD_{α} , the corresponding population means are said to be unequal. If the difference $\bar{x}_i - \bar{x}_j$ is less than LSD_{α} it is concluded that the corresponding population means are equal.

IV Student-Newman-Keuls Multiple Range Procedure (NMR)

The first procedure to use the multiple protection principle was proposed by Newman (8) working from a suggestion by Student (12). In this procedure, which was later proposed by Keuls (7), Newman made use of Student's idea for subdividing a group of ranked treatment means into subgroups asserted to be not heterogeneous and presented tables for a number of percentage points of

$$q = w/s \quad (4.1)$$

where w is the range in a sample of n treatment means ($\bar{x}_n, \bar{x}_{n-1}, \dots, \bar{x}_1$ and \bar{x}_n is the largest treatment mean whereas \bar{x}_1 is the smallest) from a normal population with variance σ^2 , so $w > 0$. s^2 is an unbiased estimate of σ^2 based on degrees of freedom. Also w and s are independent.

It is useful to find the expectation of the q given in (4.1), because

$$s^2 = \sum_{i=1}^{f+1} (\bar{x}_i - \bar{\bar{x}})^2 / f$$

where $\bar{\bar{x}}$ is an unbiased estimate of the true mean and

$$fs^2/\sigma^2 \sim \chi^2(f)$$

Hence

$$p(s) = (f^{f/2} s^{f-1} e^{-fs^2/2}) / (2^{(f-2)/2} \Gamma(f/2) \sigma^f); \quad s^2 > 0 \quad (4.2)$$

$$\begin{aligned} E(q) &= \int_0^\infty q p(q) dq \\ &= \int_0^\infty ws^{-1} p(w) p(s) dw ds; \quad \text{since } w \text{ and } s \text{ are independent} \\ &= \left[\int_0^\infty wp(w) dw \right] \left[\int_0^\infty s^{-1} p(s) ds \right] \\ &= E(w) \int_0^\infty s^{-1} p(s) ds \\ &= E(w/\sigma) \int_0^\infty \sigma s^{-1} p(s) ds \\ &= E(w/\sigma) f^{f/2} / (2^{(f-2)/2} \Gamma(f/2) \sigma^{(f-1)}) \int_0^\infty s^{f-2} e^{-fs^2/2} ds \\ &= E(w/\sigma) \sqrt{f/2} \Gamma[(f-1)/2] / \Gamma(f/2) \end{aligned} \quad (4.3)$$

The value of $E(w/\sigma)$ may be found from Tippett (1925) pp. 387-7 or from

Tables for Statisticians and Biometricians Part II, Table XXII.

If one wishes to test $H_0 (\mu_1 = \mu_2 = \dots = \mu_n / \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2)$ by means of random samples, Newman suggested using $w = q_{\alpha} s_{\bar{x}}$ as the test criterion, where $s_{\bar{x}}$ is the standard error of a mean. To apply this test, the following steps are taken:

(1) Subdivide the n treatments as prescribed by the choice of the particular set of treatment having meaning.

(2) Choose a significance level, α .

(3) Compute $s_{\bar{x}}$ and the values:

$$w_n = q_{\alpha, n} \cdot s_{\bar{x}}$$

$$w_{n-1} = q_{\alpha, n-1} \cdot s_{\bar{x}}$$

$$\vdots$$

$$w_3 = q_{\alpha, 3} \cdot s_{\bar{x}}$$

$$w_2 = q_{\alpha, 2} \cdot s_{\bar{x}} = t_{\alpha, f.d.f.} \sqrt{2} s_{\bar{x}} = \text{LSD}_{\alpha},$$

where $q_{\alpha, i}$ $i = n, n-1, \dots, 2$ can be found from tables given in Newman (8) or Federer (4).

(4) Rank the n treatment means from highest to lowest.

(5) Compare the range of n treatment means ($\bar{x}_n - \bar{x}_1$) with the calculated w_n , if the difference ($\bar{x}_n - \bar{x}_1$) is less than w_n , the process stop, and the n sample means are asserted to belong to a non-heterogeneous group. If ($\bar{x}_n - \bar{x}_1$) is greater than w_n , subdivide the means into two groups of $(n-1)$ means each, i.e. ($\bar{x}_n - \bar{x}_2$) and ($\bar{x}_{n-1} - \bar{x}_1$), and compare the range with w_{n-1} , etc. The process stops whenever the actual range of a subset is less than the calculated range. No subset of means is compared if the subset is included in a larger subset which has less than the calculated range w_i .

V Duncan's New Multiple Range Test (DMR)

Duncan (1,3) argued that at least adequate protection would be provided in any α -level procedure by securing its protection levels at the values $P_p = (1 - \alpha)^{p-1}$, where p is the number of treatments concerned in making the comparisons. The value $(1 - \alpha)^{p-1}$ is the probability of accepting the joint null hypothesis when any $(p - 1)$ orthogonal comparisons among p means are subjected to $(p - 1)$ α -level independent test. Duncan said such levels are achieved and accepted without question in sequences of independent experiments that occur in the course of an investigator's experiences throughout time.

Protection levels of this type based on degrees of freedom were used in constructing Duncan's early multiple F-test. The tables for which are available in Federer (4).

The Duncan's (1,3) multiple-range-rule (DMR) uses the protection level: $P_n = (1 - \alpha)^{p-1}$ based on degrees of freedom, where n is the total number of treatments in an experiment, and p is the number of treatments concerned in making the comparisons.

A general α -level multiple range test is given by the rule: "The difference between any two means in a set of n treatment means is significant provided that the range of each and every subset which contains the given means is significant according to an α -level range test, where $\alpha_p = 1 - \gamma_p$, $\gamma_p = (1 - \alpha)^{p-1}$, and p is the number of means in the subset concerned."

The test procedures contains the following steps:

(1) Choose the significance level, α .

(2) Compute $s_{\bar{x}}$ and R_p ,

where $s_{\bar{x}}$ is the standard error of a mean,

$R_p = q_{\alpha, p} s_{\bar{x}}$, p is the number of treatments concerned,
 α level q_p can be found from tables given in
 Fryer's (6) book.

Then:

$$R_n = q_{\alpha, n} s_{\bar{x}}$$

$$R_{n-1} = q_{\alpha, n-1} s_{\bar{x}}$$

:
:
:

$$R_2 = q_{\alpha, 2} s_{\bar{x}} = t_{\alpha, f \text{ d.f.}} 2^{1/2} s_{\bar{x}} = w_2 \quad \text{of (NMR)}$$

(3) Rank the n treatment means from the highest to the lowest denoted as $(\bar{x}_n, \bar{x}_{n-1}, \dots, \bar{x}_1)$.

(4) Compare the range of n means $(\bar{x}_n - \bar{x}_1)$ with R_n , if $(\bar{x}_n - \bar{x}_1)$ less than the calculated R_n , the process stops, and the n means are declared to be equal, otherwise the process continues. Subdivide the means into two groups $(\bar{x}_n - \bar{x}_2, \bar{x}_{n-1} - \bar{x}_1)$ each compare with R_{n-1} again, etc. The process stops whenever the actual range of a subset is less than the compute range. No subset of means is compared if the subset is included in a larger subset which is less than the calculated R_p . Note that in Duncan's DMR test, no prior F test from an analysis of variance is used, although are usually performs an analysis of variance to obtain an estimate of σ^2 , the population variance.

VI Tukey's Significant Difference Rule (TSD) Based on Allowance

In 1951 Tukey (13) introduced the concept of estimating simultaneous confidence intervals for the differences between all pairs of n treatment means. A set of intervals with a joint confidence coefficient $(1-\alpha)$ is given by

$$\{ d_{ij} - \text{TSD} < \delta_{ij} < d_{ij} + \text{TSD}, \text{ all } i, j, i \neq j \} \quad (6.1)$$

where $d_{ij} = \bar{x}_i - \bar{x}_j$, the observed treatment difference.

$\delta_{ij} = \mu_i - \mu_j$, the true treatment difference.

$\text{TSD} = q_\alpha s_{\bar{x}}$, q_α is the upper α point of the Studentized range,

$s_{\bar{x}}$ is the standard error of mean.

One of the main uses given for this procedure is in solving the multiple comparisons problem. In this use, if the confidence interval for δ_{ij} lies to the left or right to the zero, treatment i is declared significantly less or larger than treatment j . If the interval for δ_{ij} includes zero, treatment i and j are declared not significantly different. In other words, this may be expressed as the TSD (Tukey-significant-difference) rule: Any difference exceeding TSD is significant, any difference not exceeding TSD is not significant.

VII Scheffé's Significant Difference Test Based on Contrast (SSD)

The general problem, which Scheffé (11) stated, is that of making inferences about the contrasts among a set of "true means" or "true main effects" $\mu_1, \mu_2, \dots, \mu_k$ in the analysis of variance.

$$\theta = \sum_{i=1}^k c_i \mu_i \quad (7.1)$$

is called a contrast, where c_i ($i = 1, 2, \dots, k$) is a set of known constant which satisfy the condition that $\sum_{i=1}^k c_i = 0$; μ_i ($i = 1, \dots, k$) is a set of unknown parameters.

The value of the linear function for a particular set of θ will be called the value of the contrast, therefore the symbol denotes both for the contrast and the value of contrast.

We make the assumption usual in the analysis of variance that there is a set of statistics $\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_k$ and $\hat{\sigma}^2$, such that the $\hat{\mu}_i$ have a multivariate normal distribution and are statistically independent of $\hat{\sigma}^2$, that

$$\begin{aligned} E(\hat{\mu}_i) &= \mu_i, \quad (i = 1, 2, \dots, k) \\ \text{cov}(\mu_i, \mu_j) &= a_{ij} \sigma^2, \quad (i, j = 1, 2, \dots, k) \end{aligned} \quad (7.2)$$

where a_{ij} are known constants.

Using the matrix notation:

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix} \sim N \left[\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix}, \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{bmatrix} \sigma^2 \right]$$

$$\hat{\mu} \sim N[\underline{\mu}, A \sigma^2] \quad (7.3)$$

For any contrast θ , it is estimated by $\hat{\theta}$, where

$$\hat{\theta} = \sum_{i=1}^k c_i \hat{\mu}_i$$

with variance $\sigma_{\hat{\theta}}^2 = \sum_{i=1}^k \sum_{j=1}^k a_{ij} c_i c_j \sigma^2$. (7.4)

If σ^2 is unknown, it is estimated by $\hat{\sigma}^2$, and $\sigma_{\hat{\theta}}^2$ is estimated by:

$$\hat{\sigma}_{\hat{\theta}}^2 = \sum_{i=1}^k \sum_{j=1}^k a_{ij} c_i c_j \hat{\sigma}^2.$$

Now, we shall prove that $\hat{\theta}$ is an unbiased estimate of θ with variance σ^2 .

Since $\hat{\mu} \sim N[\mu, A\sigma^2]$ by (6.3);

$$\hat{\theta} = \sum_{i=1}^k c_i \hat{\mu}_i = (c_1, c_2, \dots, c_k) \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \vdots \\ \hat{\mu}_k \end{pmatrix} = C' \hat{\mu}$$

Then $C' \hat{\mu} \sim N[C' \mu, C' A C \sigma^2]$; where $C' \mu = \sum_{i=1}^k c_i \mu_i$

$$\begin{aligned} \text{and } C' A C \sigma^2 &= (c_1, c_2, \dots, c_k) \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \dots & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} \sigma^2 \\ &= \sum_{i=1}^k \sum_{j=1}^k c_i c_j a_{ij} \sigma^2. \end{aligned}$$

(Refer to Theorem 3.6 page 56 Graybill's "An Introduction to Linear Statistical Models"). Define the positive constant S from

$$S^2 = (k-1) F_{\alpha}(k-1, f) \quad (7.5)$$

where $F_{\alpha}(k-1, f)$ denotes the upper α point of the F-distribution with $k-1$ and f degrees of freedom. Then the probability is $1-\alpha$ that the value of all the contrasts simultaneously satisfy

$$\hat{\theta} - S\sigma \leq \theta \leq \hat{\theta} + S \quad (7.6)$$

no matter what the values of all unknown parameters.

This results may be used for interval estimation of all contrasts of interest. No matter how many contrasts are estimated by the method (7.6), the probability that all the statements thus made will be correct will be great than

or equal to $1 - \alpha$.

This result may also be used to declare any estimated contrast "significant different from zero or not" according to the corresponding interval (7.6) excludes $\theta = 0$ or not. More precisely, after selecting a set of c_i subject to (6.1) and thus determining a contrast, we make one of the following statements:

- (1) θ is not significantly different from zero, if $-S \hat{\sigma}_{\hat{\theta}} < \hat{\theta} < S \hat{\sigma}_{\hat{\theta}}$.
- (2) θ is significantly different from zero and positive, if $\hat{\theta} \geq S \hat{\sigma}_{\hat{\theta}}$.
- (3) θ is significantly different from zero and negative, if $\hat{\theta} \leq -S \hat{\sigma}_{\hat{\theta}}$.

As a general application, we define:

$$S^2 = [(n-1) F_{\alpha(n-1, f)}]^{1/2} \quad (7.7)$$

where n = number of treatments in the experiment, f = number of degrees of freedom associated with the error variance. $F_{\alpha, (n-1, f)}$ is the upper α point of the F-distribution.

We use $\sqrt{2} s_{\bar{x}} S$ as the criterion for testing the difference between any two means. If the difference between any two means exceeds the value of $\sqrt{2} s_{\bar{x}} S$, the two means are significantly different.

In the special case where all the $\hat{\mu}_i$ have the same variance $a_{11} \sigma^2$ and all pairs of $\hat{\mu}_i, \hat{\mu}_j$ ($i \neq j$) have the same covariance $a_{12} \sigma^2$. The Tukey's (13) method based on allowance can be applied. The probability is $1 - \alpha$ that the values of that all the contrasts simultaneously satisfy:

$$\hat{\theta} - T \hat{\sigma} \leq \theta \leq \hat{\theta} + T \hat{\sigma} \quad (7.8)$$

where the contrast T is defined as

$$T = \frac{1}{2} \sum_{i=1}^k |c_i| q (a_{11} - a_{12})^{1/2} \quad (7.9)$$

and $\hat{\sigma}$ is an estimate of σ , q is the upper α point of the Studentized range, for the range of a sample of K in the numerator, and f degrees of freedom in the denominator, this is, the upper α point of the quotient w/s , where w and

s ($s > 0$) are statistically independent, w is the range of a random sample of k standard normal deviates and $f s^2 / \sigma^2$ has the χ^2 distribution with f degrees of freedom.

VIII Example and Comments

To illustrate, consider the example given in Duncan (3), the application of a 5%-level of the above five test procedures to the results of a bread-baking experiment shown in Table 8.1. The problem of testing the differences among the mean loaf volumes for 17 different varieties of wheat. The differences involved are shown individually in Table 8.2, and the summary of difference shown in Table 8.3.

Applying the rule, the observed F-ratio is 28.35 which is much larger than $F_{.05,16,64}$ (when $\alpha = .05$ with 16 and 64 degrees of freedom).

Table 8.1 Loaf Volumes for Flour of 17 Varieties of Wheat,
Tested at Five Levels of Potassium Bromate Added in Baking

(1) Varietal Means Ranked in Order

M	P	D	C	Q	L	H	G	N	B	F	I	K	J	E	A	O
624	729	755	801	828	829	846	853	861	903	908	922	933	951	977	987	1030

(2) Analysis of Variance

Source of Variation	d.f.	Ms	F	F _{.05,16,64}
Between Varieties	16	48,559	28.35	1.81
Between Bromate Levels	4	29,134	17.01	
Error	64	1,713		

(3) Standard Error of a Varietal Mean and of a Difference

When the Number of Observations is 5

$$s_{\bar{x}} = (1713/5)^{1/2} = 18.51$$

$$s_d = 18.51 (2)^{1/2}.$$

Table 8.3 Summary of Difference Found Not-significant
by Five 5% Level Multiple Comparison Rules

Rules:	LSD	DMR	NMR	TSD	SSD
Number	34	38	49	65	69
Number Extra	-	4	11	16	4
Mangnitude of Extra Difference	-	53-55	57-76	72-94	97-99
Maximum Difference	$1.99 s_d$	$2.10 s_d$	$2.90 s_d$	$3.59 s_d$	$3.78 s_d$

Note that the SSD rule has not been proposed by Scheffé as a multiple comparison procedure for difference alone among uncorrected and homoscedastic means as presented in the above example, this was Duncan's suggestion.

8.2 Comments

In the various test procedures, different points of view can be taken about the importance of type I error relative to type II error, power and the protection level against type I error.

R. A. Fisher combined an F-test of testing if the n treatment means are equal, given equal variance, with a t-test, provided the null hypothesis is rejected. The prior F-test assumes $100 (1-\alpha) \%$ protection against type I error on the n -mean comparisons. No further effort is made to maintain any specified level of protection against type I error even for t-mean comparisons when $t < n$. But for the pairwise multiple t-test of the n treatment means comparisons, there are $\binom{n}{2}$ comparisons, without a prior F-test, the most extreme case is the comparison of the largest mean with the smallest mean. In this case, Pearson and Hartley (9,10) have shown that n treatment type I error is not 5 percent for $n > 2$, but some large value. The size of a n -treatment type I error is equal to;

$$1 - {}_f P_n (Q) \quad (8.2.1)$$

where ${}_f P_n (Q) = P_n (Q) + a_n (Q) / f + b_n (Q) / f^2$

$P_n (Q)$, $a_n (Q)$ and $b_n (Q)$ are obtained from Table I in Pearson and Hartley (10), n = number of treatments, f = degrees of freedom associated with error variance, $Q = 2^{1/2} t_{\alpha, f d/f}$

For more than two treatments in a group the size of the n treatment type I error associated with the comparison of the largest mean with the smallest mean is larger than 5 per cent if the μ_i are actually equal. For $f=40$, $t_{.05, 40} \sqrt{2}=2.86$ the n -treatment type I error associated with the comparison of the largest mean with the smallest mean is approximately equal to 27 per cent for five treatments, 59 per cent for ten treatments and 86 per cent for 20 treatments. Thus without

a prior F-test, the multiple t-test procedure for comparing the largest and the smallest means should not be used when there are more than two treatments in the experiment.

The Newman point of view was that the protection against type I error in two treatment means, three treatment means, ... , n treatment means comparisons should be kept at the same level, namely $100 (1 - \alpha) \%$. Duncan introduced in his new multiple range test with its special protection level based on degrees of freedom, the concept of protection against type I error which decreases is according to the formular $100 (1 - \alpha)^{p-1} \%$, where p is the number of treatment means concered in a subgroup.

As to Tukey's TSD and Scheffé's SSD, it can be compared under the special case where all the μ_i have the same variance $a_{11} \sigma^2$, all pairs μ_i, μ_j ($i \neq j$) have the same covariance $a_{12} \sigma^2$ by the ratio R of the squared length of the confidence interval of (7.8) and (7.6)

$$R = (S^2 \hat{\sigma}_{\hat{\theta}}^2) / (T^2 \hat{\sigma}^2) \quad (8.2.2)$$

$$\text{where } \sigma_{\hat{\theta}}^2 = (a_{11} - a_{12}) \sigma^2 \sum_{i=1}^k c_i^2 \quad (8.2.3)$$

$$\text{Because } \sigma^2 = \sigma^2 (c_1, c_2, \dots, c_k) \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}$$

$$\begin{aligned} &= \sigma^2 [c_1^2 a_{11} + c_1 (c_2 + c_3 + \dots + c_k) a_{12} + \dots + \\ &\quad c_2^2 a_{11} + \dots + c_k^2 a_{11} + c_k (c_1 + \dots + c_{k-1}) a_{12}] \\ &= \sigma^2 [a_{11} \sum_{i=1}^k c_i^2 + a_{12} (c_1 (c_1 + \dots + c_k - c_1) + \\ &\quad + \dots + c_k (c_1 + \dots + c_k - c_k))] \\ &= \sigma^2 [a_{11} \sum_{i=1}^k c_i^2 + a_{12} (c_1 + \dots + c_k)^2 - a_{12} (c_1^2 + \\ &\quad + \dots + c_k^2)] \end{aligned}$$

$$\begin{aligned}
&= \sigma^2 \left[a_{11} \sum_{i=1}^k c_i^2 + a_{12} \left(\sum_{i=1}^k c_i \right)^2 - a_{12} \sum_{i=1}^k c_i^2 \right] \\
&= \sigma^2 \left[a_{11} \sum_{i=1}^k c_i^2 - a_{12} \sum_{i=1}^k c_i^2 \right] \\
&= \sigma^2 (a_{11} - a_{12}) \sum_{i=1}^k c_i^2
\end{aligned}$$

and $T = \left[\sum_{i=1}^k |c_i| q (a_{11} - a_{12})^{1/2} \right] / 2$ by (7.9). $\hat{\sigma}_\delta^2$ is an estimate of σ_δ^2 . Therefore

$$R = (S^2 / q^2) \left[\sum_{i=1}^k c_i^2 / \left(\sum_{i=1}^k |c_i| / 2 \right)^{1/2} \right]. \quad (8.2.4)$$

Table given in Scheffe (11) shows how the relative efficiency of the two methods varies with k . As k increases, equation (7.6) gets relatively worse on the difference of $\mu_i - \mu_j$. If we are interested in testing the pairwised means, i.e. some μ_i are equal, some are not, we should use Tukey's TSD method, because it gives a shorter confidence interval.

IX Bayesian Approach to Multiple Comparisons

9.1 Before we start to discuss the Bayesian approach to multiple comparisons, it may be helpful at this point to review very briefly the notations, definitions and technique from the general decision theory due to Wald (14).

Consider the problem of choosing between two decisions

$$\begin{aligned} d_0 &: \text{accept } H_0 : \tau = \tau_0, \\ d_1 &: \text{accept } H_1 : \tau = \tau_1. \end{aligned} \quad (9.1.1)$$

Let the loss function for the decision d_0 and d_1 be represented by

$$L_0(\tau) = \begin{cases} 0, & \tau = \tau_0, \\ c_0, & \tau = \tau_1, \end{cases} \quad \text{and} \quad L_1(\tau) = \begin{cases} c_1, & \tau = \tau_0, \\ 0, & \tau = \tau_1, \end{cases} \quad (9.1.2)$$

where c_0 and c_1 are positive numbers with $c_0 > c_1 > 0$.

Let prior probabilities for H_0 and H_1 be represented by

$$P(\tau) = \begin{cases} p_0, & \tau = \tau_0, \\ p_1, & \tau = \tau_1, \end{cases} \quad (9.1.3)$$

where $p_0 + p_1 = 1$.

We will use an indicator function to represent a two decision test rule

$$\phi(t) = \begin{cases} 0, & t \in R_0, \\ 1, & t \in R_1, \end{cases} \quad (9.1.4)$$

where R_0 and R_1 are the sets of t values for which the corresponding decision are to be made.

The value of 0 or 1 for $\phi(t)$ at any t thus indicates that the decision d_0 or d_1 is to be made at that t . The probabilities of d_0 and d_1 for any τ will be denoted by $P[d_i|\tau]$, $i = 0, 1$. Similarly, in representing a multidecision rule with p decisions d_1, \dots, d_p we will use

$$\phi_i(t) = \begin{cases} 0, & t \notin R_i, \\ 1, & t \in R_i, \end{cases} \quad (9.1.5)$$

where R_1 is the set of all points t at which the corresponding decision is to be made.

The risk of any decision rule $\emptyset(t) = [\emptyset_1(t), \emptyset_2(t), \dots, \emptyset_p(t)]$ for any p decisions d_1, d_2, \dots, d_p at any parameter point is the expected loss

$$R(\emptyset | \tau) = \sum_{i=1}^p L_i(\tau) P[d_i | \tau] \quad (9.1.6)$$

where $L_i(\tau)$ are the loss for d_i , $i = 1, 2, \dots, p$ at τ .

The average risk for any p -decision rule $\emptyset(t)$ as above is the weighted average

$$A(\emptyset) = \int_{-\infty}^{\infty} R(\emptyset | \tau) P(\tau) d\tau, \quad (9.1.7)$$

where $P(\tau)$ is the prior density for τ .

A Bayes rule $\emptyset(t)$ for any problem is the rule which minimizes the average risk (9.1.7).

9.2 With much help from the more general work of Lehmann^{*}, it is possible to solve the multiple comparisons problem by Bayesian approach, Duncan (2,3) developed the Bayes solution for the multiple comparisons problems to a tolerated error probability "of the first kind" for each simple difference, that is independent of the number of treatments being compared.

There are two common types of Student- t problems closely related to the Bayesian multiple comparisons problem which will first be discussed.

(a) A two-decision Student- t problem. Given a random observation t from a noncentral t -distribution with non-centrality parameter τ and v degrees of freedom, a common problem is that of choosing between the two decisions

$$d_0 : \text{decide } \tau < \Delta \quad \text{and} \quad d_1 : \text{decide } \tau > \Delta,$$

^{*} Refer to Lehmann "A theory of some multiple decision problems I, II" Ann. Math. Stat., 21, 1-26 and Ann. Math. Stat., 28, 1-25, 1957.

where $\Delta > 0$.

In the theory of hypothesis testing, this problem is often more loosely regarded as that of testing $H_0 : \tau \leq 0$ against $H_1 : \tau > 0$.

Let the loss function be

$$\begin{aligned} L_0(\tau) &= \begin{cases} 0, & \tau \leq 0, \\ k_0 \tau, & \tau > 0, \end{cases} \\ L_1(\tau) &= \begin{cases} k_1 |\tau|, & \tau \leq 0, \\ 0, & \tau > 0, \end{cases} \end{aligned} \quad (9.2a.1)$$

where k_0 and k_1 are positive constants such that $k_1 > k_0$ and τ has a normal prior density $P(\tau)$ with mean zero and variance γ^2 , then the Bayes rule for this problem is of the common form

$$\phi_*(t) = \begin{cases} 0, & t < t_*, \\ 1, & t > t_*, \end{cases} \quad (9.2a.2)$$

where $t_* = t_*(k, v, \gamma^2)$ is a significant t value which is determined by the ratio $k = k_1 / k_0$ from loss function, the degrees of freedom v for t and the variance γ^2 of the prior density for τ . The significant t values for different combinations of k, v, γ^2 are given in Table 9.2a.1. We will call k, v, γ^2 the error seriousness ratio, the error degrees of freedom and the risk-weighting variance ratio respectively. It is of interest to note that a loss ratio of 100 ($\log k = 2$) infinite error degrees of freedom ($v = \infty$) a risk-weighting variance ratio of 3 ($\gamma^2 = 3$) gives a t_* of 1.987 close to that 1.960 of a .025 level Student- t test of $H_0 : \tau \leq 0$.

Table 9.2a.1 Minimum-Average-Risk Significant t Values (t_{α} Values)

		Log k							
γ^2	v	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5
∞	1	0.0	0.375	0.807	1.353	2.102	3.160	4.685	6.854
	2	0.0	0.413	0.860	1.379	2.012	2.814	3.851	5.208
	4	0.0	0.434	0.884	1.367	1.900	2.502	3.197	4.010
	6	0.0	0.443	0.891	1.356	1.848	2.374	2.948	3.580
	14	0.0	0.451	0.898	1.340	1.779	2.217	2.654	3.099
	∞	0.0	0.457	0.902	1.326	1.721	2.091	2.436	2.759
3	1	0.0	0.444	1.053	2.053	∞	∞	∞	∞
	2	0.0	0.484	1.060	1.926	4.077	∞	∞	∞
	4	0.0	0.506	1.056	1.718	2.623	4.178	9.595	∞
	6	0.0	0.515	1.053	1.653	2.370	3.308	4.732	7.706
	14	0.0	0.522	1.047	1.528	2.136	2.724	3.360	4.074
	∞	0.0	0.528	1.041	1.531	1.987	2.414	2.813	3.186
1	1	0.0	0.572	1.930	∞	∞	∞	∞	∞
	2	0.0	0.610	1.532	8.741	∞	∞	∞	∞
	4	0.0	0.629	1.395	2.648	8.592	∞	∞	∞
	6	0.0	0.637	1.353	2.303	3.980	13.625	∞	∞
	14	0.0	0.642	1.308	2.030	2.859	3.891	5.326	7.818
	∞	0.0	0.646	1.275	1.875	2.433	2.957	3.445	3.902
0.5	1	0.0	0.767	∞	∞	∞	∞	∞	∞
	2	0.0	0.785	2.292	∞	∞	∞	∞	∞
	4	0.0	0.791	1.963	9.243	∞	∞	∞	∞
	6	0.0	0.794	1.800	3.777	∞	∞	∞	∞
	14	0.0	0.792	1.653	2.693	4.162	6.670	∞	∞
	∞	0.0	0.792	1.562	2.296	2.980	3.622	4.219	4.779

(b) A related three-decision Student-t problem. Given a similar observed t value, a problem related to that of (a) is one of choosing between the three common decisions,

$$\begin{aligned} d_0 : & \text{decide } \tau \leq \Delta, \\ d_1 : & \text{decide } \tau > \Delta, \\ \text{and } d_2 : & \text{decide } \tau < -\Delta, \end{aligned} \quad (9.2b.1)$$

where, as before, $\Delta > 0$. In the language of the experimenter d_0 , d_1 and d_2 are the decisions that τ is not significantly different from zero, that τ is significantly greater than zero and that τ is significantly less than zero, respectively. Again, we assume the loss function to be a simple linear form,

$$\begin{aligned} L_0^{(2)}(\tau) &= \begin{cases} 0, & \tau = 0, \\ c_0|\tau|, & \tau \neq 0, \end{cases} \\ L_1^{(2)}(\tau) &= \begin{cases} c_1|\tau|, & \tau \leq 0, \\ 0, & \tau > 0, \end{cases} \\ L_2^{(2)}(\tau) &= \begin{cases} 0, & \tau > 0, \\ c_1\tau, & \tau \leq 0, \end{cases} \end{aligned} \quad (9.2b.2)$$

where c_0 and c_1 are positive constants such that $c_1 - c_0 > c_0$ and the prior density for τ be the same normal as defined in (a). Then the Bayes rule for this problem is obtained as follows:

First of all, we consider that the three-decision subset system

$$w_0 : |\tau| \leq \Delta, \quad w_1 : \tau > \Delta, \quad w_2 : \tau < -\Delta, \quad (9.2b.3)$$

can be expressed as the restricted product of two component two-decision subset systems like that of the previous problem in (a), namely

$$\text{Component system for } +\tau : w_0^+ : \tau < \Delta, \quad w_1^+ : \tau > \Delta,$$

$$\text{Component system for } -\tau : w_0^- : -\tau < \Delta, \quad w_1^- : -\tau > \Delta.$$

(9.2b.4)

Thus

$$w_0 = w_0^+ \cap w_1^-, \quad w_1 = w_1^+ \cap w_1^-, \quad w_2 = w_0^+ \cap w_1^+. \quad (9.2b.5)$$

The intersection of $w_1^+ \cap w_1^- = \emptyset$ hence is excluded. Each of the main decisions is equivalent to two joint component decisions

$$\begin{aligned} d_0 & \text{ to } d_0^+ \text{ with } d_0^-, \\ d_1 & \text{ to } d_1^+ \text{ with } d_0^-, \\ d_2 & \text{ to } d_0^+ \text{ with } d_1^-, \end{aligned} \quad (9.2b.6)$$

where d_i^α is the decision $\tau \in w_i^\alpha$, $\alpha = +, -, i = 0, 1$.

Second, if we put $k_1 = c_1 - c_0$ and $k_0 = c_0$, the loss for the main decisions can be expressed as the sums of losses for its component decision given by the two-decision loss function (9.2a.1) in (a). Any decisions rule $\emptyset^{(2)}(t)$ for the three-decision problem can also be expressed in terms of two component two-decision rules. For this purpose it is convenient to first reexpress the two-decision function $\emptyset(t)$ in the two-element vector form

$$\emptyset(t) = [\emptyset_0(t) \quad \emptyset_1(t)],$$

where $\emptyset_0(t) = 1 - \emptyset(t)$ and $\emptyset_1(t) = \emptyset(t)$. In this form the bayes rule $\emptyset(t)$ of (a) appears as

$$\emptyset_*(t) = \begin{cases} (1 & 0), & t < t_*, \\ (0 & 1), & t > t_*. \end{cases} \quad (9.2b.7)$$

With this vector notation for the two-decision function and the components of the three decisions, we can write

$$\begin{aligned} \emptyset^{(2)}(t) &= [\emptyset_0^{(2)}(t) \quad \emptyset_1^{(2)}(t) \quad \emptyset_2^{(2)}(t)] \\ &= [\emptyset_0^+(t) \emptyset_0^-(t) \quad \emptyset_1^+(t) \emptyset_0^-(t) \quad \emptyset_0^+(t) \emptyset_1^-(t)], \end{aligned} \quad (9.2b.8)$$

where $\emptyset_i^\alpha(t) = 0$, or 1 denotes the not making or making of decision d_i^α , $\alpha = +, -, i = 0, 1$.

From the result of (a), the component solutions are

$$\phi_{*}^{+}(t) = \begin{cases} (1 & 0), & t < t_{*}, \\ (0 & 1), & t > t_{*}, \end{cases} \quad (9.2b.9)$$

$$\phi_{*}^{-}(t) = \begin{cases} (1 & 0), & -t < t_{*}, \\ (0 & 1), & -t > t_{*}. \end{cases}$$

Since $\phi_{1*}^{+}(t) = 1$ only if $t > t_{*}$, $\phi_{1*}^{-}(t) = 1$ only if $t > -t_{*}$, and t_{*} is positive (from $k_1 > k_0$) we have $\phi_{1*}^{+}(t) \phi_{1*}^{-}(t) = 0$ for all t .

The required Bayes rule is given by

$$\begin{aligned} \phi_{*}^{(2)}(t) &= [\phi_{0*}^{+}(t) \phi_{0*}^{-}(t) \quad \phi_{1*}^{+}(t) \phi_{0*}^{-}(t) \quad \phi_{0*}^{+}(t) \phi_{1*}^{-}(t)] \\ &= \begin{cases} (1 & 0 & 0), & (t < t_{*}) (t > -t_{*}) = t < t_{*}, \\ (0 & 1 & 0), & (t > t_{*}) (t > -t_{*}) = t > t_{*}, \\ (0 & 0 & 1), & (t < t_{*}) (t < -t_{*}) = t < -t_{*}, \end{cases} \quad (9.2b.10) \end{aligned}$$

where the significant t ratio $t_{*} = t_{*}(k, v, \gamma^2)$ is the same as that of (a) with the loss ratio now given by $k = (c_1 / c_0) - 1$, and where $\phi_{i}^{(2)}(t) = 0$ or 1 denotes the not making or making of the decision d_i , $i = 0, 1, 2$.

(c) Asymmetric multiple comparisons problem. Given $N = n(n-1)/2$ t statistics of the form

$$t_{pq} = (\hat{\mu}_p - \hat{\mu}_q) / S_{\hat{\mu}_p - \hat{\mu}_q}, \quad pq \in N, \quad (9.2c.1)$$

with non-centrality parameters of the form

$$\tau_{pq} = (\mu_p - \mu_q) / \sigma_{\hat{\mu}_p - \hat{\mu}_q}, \quad pq \in N, \quad (9.2c.2)$$

where N denotes the set of pairs $\{1, 2; 1, 3; \dots; (n-1), n\}$.

A common multiple comparisons problem is that of choosing between the three decisions

$$\begin{aligned} d_{pq}^0 : & \quad \tau_{pq} \in w_{pq}^0, \\ d_{pq}^1 : & \quad \tau_{pq} \in w_{pq}^1, \\ d_{pq}^2 : & \quad \tau_{pq} \in w_{pq}^2, \end{aligned} \quad (9.2c.3)$$

simultaneously for all $pq \in N$, where

$$w_{pq}^0 : |\tau_{pq}| \leq \Delta, \quad w_{pq}^1 : \tau_{pq} > \Delta, \quad w_{pq}^2 : \tau_{pq} < -\Delta. \quad (9.2c.4)$$

Under a generalizations of the linear additive loss functions and normal prior density used as before, the losses are defined as the sum of losses (9.2b.2) for each of the component decisions involved; that is

$$L_i^{(n)}(\tau) = L^{(n)}(\tau, d_i) = \sum_{pq \in N} L_{j_{pq}}^{(2)}(\tau_{pq}), \quad j_{pq}^1 = 0, 1, 2; \quad (9.2c.5)$$

$$i = 0, 1, \dots, M-1,$$

where $M = N_0$. of multiple comparisons subsets. (for example, if $n=3$, $M=19$).

From the additive losses assumption it follows that the average risk for any decision rule $\theta^{(n)}(t) = [\theta_0^{(n)}(t) \dots \theta_{M-1}^{(n)}(t)]$ may be expressed as the sum of average risks for component three-decision rules

$$\theta^{pq}(t) = [\theta_0^{pq}(t) \quad \theta_1^{pq}(t) \quad \theta_2^{pq}(t)], \quad (9.2c.6)$$

provided again that the component rules are compatible. Thus we can write

$$A(\xi_n, \theta^{(n)}) = \sum_{pq \in N} A(\xi_n, \theta^{pq}(t)). \quad (9.2c.7)$$

As before, the Bayes rule

$$\theta_{*}^{(n)}(t) = [\theta_{0*}^{(n)}(t) \dots \theta_{M-1*}^{(n)}(t)],$$

for the multiple comparisons problem is formed by the products

$$\theta_{i*}^{(n)}(t) = \prod_{pq \in N} \theta_{j_{pq}}^{pq}(t), \quad j_{pq}^1 = 0, 1, 2, \quad (9.2c.8)$$

of the elements of the Bayes rules $\theta_{*}^{(n)}(t)$ minimizing

$$A(\xi_n, \theta^{pq}(t)), \quad pq \in N, \quad (9.2c.9)$$

provided these are compatible. This is minimized by

$$\theta_{*}^{pq}(t) = \begin{cases} (1 \ 0 \ 0), & (\tau_{pq}' < t_{*}) = (\tau_{pq}' < t_{*}/R), \\ (0 \ 1 \ 0), & (\tau_{pq}' > t_{*}) = (\tau_{pq}' > t_{*}/R), \\ (0 \ 0 \ 1), & (\tau_{pq}' < -t_{*}) = (\tau_{pq}' < -t_{*}/R), \end{cases} \quad (9.2c.10)$$

where $t'_{pq} = R t_{pq}$ for all $pq \in N$, and $R^2 = v' / [v + \sum_{i=2}^{n-1} t^2 / (1+\gamma^2)]$,
 $v' = v + n - 2$.

The Bayes rule for the multiple comparisons problem is given by the simultaneous application (9.2c.8) of all $N = n(n-1)/2$ of the three-decision Bayes rules (9.2c.10).

9.3 Example of Bayes rule:

Suppose that n samples of yields have been obtained for n new treatments. For each and every pair (a,b) of treatments it is required to decide whether a can be recommended as the superior, whether b can or whether to withhold recommendations on both. The loss ratio c being the same for all pairs, and are additive in giving the losses for each of the joint decisions to which they contribute. Risks are to be averaged with respect to a normal independent prior density for each of the means $\mu_1, \mu_2, \dots, \mu_n$ each with the same mean and same variance $\gamma^2 \sigma^2 / r$.

The required decision rule is then given by the simultaneous application (9.2c.8) of (9.2c.10) where $t_* = t_*(k, v', \gamma^2)$ with $k = c-1$ and $v' = v + n - 2 = n(r-1) + n - 2$ and where t'_{pq} can be obtained by analysis-of-variance type steps as follow: Put S_t, S_{pq}, S'_{pq} and S_e for the treatment sum of squares, the sum of squares for the pq difference, the residual sum of squares for the pq difference and the error sum of squares

$$S_t = r \sum_{i=1}^n \bar{x}_i^2 - C, \quad S_{pq} = r (\bar{x}_p - \bar{x}_q)^2 / 2,$$

$$S'_{pq} = S_t - S_{pq}, \quad S_e = \left[\sum_{i=1}^n \sum_{j=1}^r x_{ij} - C \right] - S_t,$$

respectively, where C is the correction term $(\sum_{i=1}^n \sum_{j=1}^r x_{ij})^2 / nr$.

Let $s_{pq}^{'2}$ denote the pooled estimate of σ^2 obtained as

$$s_{pq}^{'2} = [S_e + S_{pq}' / (1 + \gamma^2)] / v'.$$

Then t_{pq}' may be obtained as the square root of the variance ratio

$$t_{pq}^{'2} = S_{pq} / s_{pq}^{'2},$$

t_{pq}' is given the same sign as $\bar{x}_p - \bar{x}_q$.

A more convenient rule for application can be obtained by expressing the inequalities $t_{pq}^{'2} \lesseqgtr t_*^2$ in the form $d_{pq}^2 \lesseqgtr d_*^2$ where d_* is a least significant value for the difference $d_{pq} = \bar{x}_p - \bar{x}_q$. From $t_{pq}^{'2} = t_*^2$ we get

$$t_*^2 = S_{pq} / s_{pq}^{'2} = v' S_{pq} / [S_e + (S_t - S_{pq}) / (1 + \gamma^2)]$$

$$t_*^2 [S_e + (S_t - S_{pq}) / (1 + \gamma^2)] = v' S_{pq}$$

$$S_{pq} [v' + t_*^2 / (1 + \gamma^2)] = t_*^2 [S_e + S_t / (1 + \gamma^2)].$$

But $S_{pq} = r d_{pq}^2 / 2$, hence this gives $d_{pq}^2 = d_*^2$ where

$$d_* = \left\{ \frac{2}{r} t_*^2 [S_e + S_t / (1 + \gamma^2)] / [v' + t_*^2 / (1 + \gamma^2)] \right\}^{1/2}.$$

From this and a check on signs it follows that the multiple comparisons Bayes rule is given by the simultaneous application of the rules

$$\phi_*^{pq}(t) = \begin{cases} (100), & |d_{pq}| < d_*, \\ (010), & d_{pq} > d_*, \\ (001), & d_{pq} < -d_*. \end{cases}$$

Note that in the complete absence of prior criteria for choosing γ^2 , we can obtain an estimate of it from the variance ratio $F = S_t / (n-1) s^2$, where s^2 is the pooled within-sample variance estimate

$$s^2 = \sum_{i=1}^k \sum_{j=1}^k (x_{ij} - \bar{x}_i)^2 / n(r-1).$$

As for choosing an error-weight ratio k , a useful approach will be to choose a k ratio which would tend to give the same order of t value as would have seemed desirable in a corresponding two treatment-experiment situation in the past.

X Monte Carlo Studies of Fisher's LSD

Since many comments have been made about protection against type I error and power under various test rules, Monte Carlo studies of Fisher's LSD under a completely randomized design experiment were performed using digital computers.

Although other definitions are possible, herein one defines the protection against type I error by the ratio:

$$\text{Protection} = \frac{100 \text{ (number of correct 2-mean decisions when } \mu_1 = \mu_j \text{)}}{\text{Total number of 2-mean decisions when } \mu_1 = \mu_j}$$

One defines the power of the test by

$$\text{Power} = \frac{100 \text{ (number of correct decisions when } \mu_1 \neq \mu_j \text{)}}{\text{Total number of 2-mean decisions when } \mu_1 \neq \mu_j}$$

Following are some illustrations relevant to table 10.1 with sample size equal 10 observations in all instances.

(1) Sampling of $N(5,4)$, $N(5,4)$, $N(5,4)$, $N(7,4)$ and $N(7,4)$ simultaneously, denoted as $N(5,5,5,7,7;4)$. In this experiment, there are $\binom{5}{2} = 10$ decisions to be made about the null hypothesis $H_0 : (\mu_1 = \mu_j)$ for $i, j = 1, 2, 3, 4, 5$, $i \neq j$, given equal variance. It is known that 4 of the 10 decisions should be decisions to accept H_0 , and 6 of them should be decisions to reject it because one knows the true population mean. Hence after a sufficient number of sets of samples, it is possible to obtain an estimate of the probabilities of type I error and type II error on a per-decision basis as defined above. After simulations of 672 sets of samples, the protection was 96 per cent when $\alpha = .05$ and the power was 52 per cent.

(2) Sampling of $N(5,4)$, $N(5,4)$, $N(5.5,4)$, \dots , $N(7,4)$ simultaneously is denoted by $N(5,5,5.5,5.5,6,6,6.5,6.5,7,7; 4)$. There are $\binom{10}{2} = 45$ possible decisions in testing the pairwise equality of the population means. Because one

knows true means of the population, it is clear that there are exactly 5 cases for which $\mu_i = \mu_j$ for all $i \neq j$, and 40 for which $\mu_i = \mu_j$ for all $i \neq j$, after the simulation it was found that the protection against type I error was 96 per cent for $\alpha = .05$ and the power was 18 per cent.

(3) Sampling of $N(6,6)$, $N(7,6)$, $N(8,6)$, $N(9,6)$, and $N(10,6)$ simultaneously denoted by $N(6,7,8,9,10; 6)$. None of the true means are equal. After 200 sets of samples we found the power of the test is 40 per cent and 51 per cent for $\alpha = .05$ and $\alpha = .10$ respectively.

The results shown in table 10.1 gives some interesting evidence regarding the comparative powers of these to detect the true population difference between means under what could represent practical conditions. From the table, there is no support for the fear that Fisher's LSD test lacks protection against type I error. As expected, the power of the test depends on the actual differences which exist among the true means μ_i and also depends on the size of the population variance.

Table 10.1 Results from a Monte Carlo Study of Fisher's LSD Method of Making Multiple Comparisons for a Type I Error Experiment

Experimental Situation	Correct Decision % (NO ^g in Paratheses are NO ^g of Decisions)							
	When $\mu_1 = \mu_j$ (protection)				When $\mu_1 \neq \mu_j$ (power)			
	$\alpha = .05$		$\alpha = .10$		$\alpha = .05$		$\alpha = .10$	
N[5,5,5,7,7; 4]	(2688)	96	(848)	90	(4032)	52	(1272)	69
; 6]	(800)	96	(800)	90	(1200)	39	(1200)	53
; 8]	(800)	95	(800)	91	(1200)	26	(1200)	37
;10]	(800)	96	(800)	92	(1200)	19	(1200)	30
;12]	(800)	96	(800)	92	(1200)	15	(1200)	24
;14]	(800)	96	(800)	93	(1200)	14	(1200)	22
;16]	(1928)	97	(800)	93	(2892)	13	(1200)	17
N[5,5,5,5,5,5,								
7,7,7,7; 4]	(11298)	95	(4263)	89	(12912)	57	(4872)	71
; 6]	(4200)	96	(4200)	90	(4800)	38	(4800)	53
; 8]	(4200)	96	(4200)	92	(4800)	27	(4800)	39
;10]	(4200)	97	(4200)	93	(4800)	20	(4800)	31
;12]	(4620)	96	(4620)	92	(5280)	17	(5280)	25
;14]	(4725)	97	(4725)	93	(5400)	14	(5400)	23
;16]	(6300)	97	(4200)	93	(7200)	12	(4800)	21
N[5,5,5,5,5,6,6,								
6,5,6,5,7,7; 4]	(1000)	96	(1000)	92	(8000)	20	(8000)	29
; 6]	(1000)	96	(1000)	92	(8000)	13	(8000)	22
; 8]	(1000)	97	(1000)	93	(8000)	9	(8000)	15
;10]	(1000)	96	(1000)	93	(8000)	8	(8000)	13
;12]	(1000)	96	(1000)	93	(8000)	7	(8000)	11
;14]	(1000)	96	(1000)	93	(8000)	6	(8000)	10
;16]	(1000)	98	(1000)	96	(8000)	4	(8000)	8
N[5,5,6,6,6,7,7,								
8,8,9,9; 4]	(1000)	95	(1000)	90	(8000)	56	(8000)	64
; 6]	(1000)	96	(1000)	91	(8000)	44	(8000)	53
; 8]	(1000)	95	(1000)	91	(8000)	34	(8000)	45
;10]	(1000)	95	(1000)	89	(8000)	31	(8000)	40
;12]	(1000)	95	(1000)	90	(8000)	25	(8000)	34
;14]	(1000)	95	(1000)	91	(8000)	23	(8000)	32
;16]	(1000)	94	(1000)	90	(8000)	20	(8000)	30

Correct Decision %

(NO₂ in parentheses are NO₂ of Decisions)

Experimental Situation	When $\mu_1 = \mu_j$ (protection)		When $\mu_1 = \mu_j$ (power)	
	$\alpha = .05$	$\alpha = .10$	$\alpha = .05$	$\alpha = .10$
N[5,5,5,5,5,5,5,5,5, 5.5,6,6,6,6,6,5,6.5, 6.5,6.5,7,7,7,7; 4]	(4500)	95 (3000)	92 (16000)	23 (10000) 30
; 6]	(6000)	96 (6000)	92 (32000)	13 (32000) 21
; 8]	(6000)	96 (6000)	92 (32000)	10 (32000) 18
;10]	(5250)	97 (5250)	93 (28000)	7 (28000) 15
;12]	(3000)	97 (3000)	93 (16000)	6 (16000) 12
;14]	(3000)	97 (3000)	92 (16000)	6 (16000) 10
;16]	(6000)	98 (6000)	95 (32000)	5 (32000) 9
N[5,6,7,8,9; 4]	Not Applicable		(9030)	52 (2000) 61
; 6]	"		(2000)	41 (2000) 51
; 8]	"		(2000)	35 (2000) 44
;10]	"		(2000)	25 (2000) 34
;12]	"		(2000)	23 (2000) 32
;14]	"		(2000)	20 (2000) 29
;16]	"		(6510)	16 (2000) 26
N[5,5,7,7,9,9, 11,11,13,13; 4]	(1000)	95 (1000)	91 (8000)	85 (8000) 90
; 6]	(1000)	96 (1000)	90 (8000)	76 (8000) 82
; 8]	(1000)	96 (1000)	92 (8000)	72 (8000) 78
;10]	(1000)	96 (1000)	90 (8000)	65 (8000) 73
;12]	(1000)	94 (1000)	89 (8000)	61 (8000) 69
;14]	(1000)	95 (1000)	90 (8000)	57 (8000) 65
;16]	(1000)	95 (1000)	90 (8000)	54 (8000) 62
N[5,5,5,5,5,5, 5,5,5,8,8,8,8, 8,8,8,8,8,8; 4]	(9000)	94 (9000)	90 (10000)	92 (10000) 96
; 6]	(17550)	95 (17550)	90 (19500)	78 (19500) 86
; 8]	(13500)	95 (13500)	89 (15000)	65 (15000) 76
;10]	(17640)	95 (17640)	90 (19600)	55 (19600) 67
;12]	(9000)	95 (9000)	90 (10000)	49 (10000) 63
;14]	(9000)	95 (9000)	90 (10000)	45 (10000) 58
;16]	(16380)	95 (16380)	90 (18200)	37 (18200) 50

Correct Decision %
(NO^s in paratheses are NO^s of Decisions)

Experimental Situation	When $\mu_1 = \mu_j$ (protection)		When $\mu_1 \neq \mu_j$ (power)			
	$\alpha = .05$	$\alpha = .10$	$\alpha = .05$		$\alpha = .10$	
N[5,7,9,11,13; 4]	Not Applicable		(5960)	77	(2000)	88
; 6]	"		(2000)	77	(2000)	82
; 8]	"		(2000)	70	(2000)	76
;10]	"		(2000)	68	(2000)	73
;12]	"		(2000)	62	(2000)	70
;14]	"		(2000)	56	(2000)	65
;16]	"		(2000)	50	(2000)	63
N[5,5,5,6,6.5,7,7.5, 8,8.5,9,9.5; 4]	Not Applicable		(16920)	48	(8100)	57
; 6]	"		(9000)	41	(9000)	50
; 8]	"		(9000)	34	(9000)	42
; 10]	"		(9000)	28	(9000)	38
; 12]	"		(9000)	23	(9000)	33
; 14]	"		(9000)	19	(9000)	28
; 16]	"		(9000)	18	(9000)	26

XI Acknowledgement

The writer wants to express her sincere appreciation to her major professor, Dr. Holly C. Fryer, for suggesting this topics and for his advice and assistance during the parperation of this report.

XII Reference

- (1) Duncan, D. B. " Multiple range and multiple F-test ", Biometrics 11, 1955, 1-42.
- (2) Duncan, D. B. " Bayes rules for a common multiple comparison problems and related Student-t problems ", Ann. Math. Stat. 32, 1961, 1013-1033.
- (3) Duncan, D. B. " A Bayesian approach to multiple comparison " Technometrics, vol. 7, NO. 2, 1965, 171-222.
- (4) Federer, W. T. " Experimental Design -- Theory & Application ", 1955, The MacMillan CO., New York.
- (5) Fisher, R. A. " The Design of Experiments ", Sixth eds., London, Oliver & Boyd, 1951.
- (6) Fryer, H. C. " Concepts & Methods of Experimental Statistics ", Allyn & Bacon, 1966.
- (7) Keuls, M. " The use of the Studentized range in connection with an analysis of variance ", Euphytica 1, 1952, 112-122.
- (8) Newman, D. " The distribution of the range in samples from a normal population, expressed in terms of an independent estimate of standard deviation ", Biometrics 5, 1939, 99-114.
- (9) Pearson, E. S. and Hartley, H. O. " The probability integral of the range in samples of n observations from normal population ", Biometrika 32, 1942, 301-310.
- (10) Pearson, E. S. and Hartley, H. O. " Tables of the probability integral of the Studentized range ", Biometrika 33, 89-99.
- (11) Scheffe, H. " A method for judging all contrasts in the analysis of variance ", Biometrika 40, 1953, 87-104.
- (12) Student, " Error of routine analysis ", Biometrika 19, 1927, 151-164.
- (13) Tukey, J. W. " Quick and dirty methods in statistics, Part II, simple analysis for standard designs ", Proc. Fifth Ann. Convention, American Society for Quality Control, 189-197, 1951.
- (14) Wald, A. " Statistical Decision Functions ", John Wiley, New York, 1950.

THE MULTIPLE COMPARISON PROCEDURES

by

NANCY CHIANG

B. A. NATIONAL TAIWAN UNIVERSITY, 1958

AN ABSTRACT OF A MASTER REPORT

submitted in partial fulfillment of the

requirement for the degree

MASTER OF SCIENCE

Department of Statistics

Kansas State University

Manhattan, Kansas

1967

The main concern of this report is the symmetric multiple comparisons problem of simultaneously testing the difference between several sample (treatment) means taken in all possible pairs. Various test procedures for solving this problem are discussed and illustrated. These include: Fisher's LSD (Least Significant Difference Rule), which is used whenever the observed F-ratio is greater than the corresponding tabulated F value at a predetermined α level of significance in an analysis of variance setting; Student-Newman-Keuls NMR (Multiple Range Test) and Duncan's DMR (New Multiple Range Test), which are based on multiple protection principle, instead of working on $\binom{n}{2}$ pairwise comparison for all possible pairs, compare the range of the treatment means with the test statistics; Tukey's TSD (Significant Difference Rule), using the concept of estimating simultaneous confidence intervals for the difference between all pairs of n treatment means; Scheffe's SSD (Significant Difference Rule) based on contrast, which is a linear function of a set of "true means" or "true main effects" such that the summation over all the known constant is equal to zero, use the test statistic to declare any estimated contrast significant greater than zero, less than zero or not significantly from zero; recently, Duncan developed a Bayesian approach to the multiple comparison problem, based on the minimum average risk, starting from two related Student-t problems, the Bayesian comparisons rule follows. A Monte Carlo estimation of protection and power of Fisher's LSD test has been performed on an IBM 1410 computer, a complete set of its results is also presented.